2 THE HUMAN-MACHINE INTERFACE CHALLENGE

Gregory Francis
Clarence E. Rash
Michael B. Russo

In Chapter 1, *The Military Operational Environment*, the military user of helmet-mounted displays (HMDs) and operational environment were described in a general fashion. In this chapter, we try to identify the essential issues that must be considered when attempting to use an HMD or head-up display (HUD). We begin by suggesting that the human-machine interface (HMI) challenge of HMD design is to *use robust technology to organize and present information in a way that meets the expectations and abilities of the user*. This chapter outlines some of the main concepts that are relevant to this challenge. Subsequent chapters describe important details about human sensory, perceptual, and cognitive systems and describe the characteristics, abilities, and limitations of HMD systems. Additional engineering-related information about HMDs can be found in this book's predecessor, *Helmet-Mounted Displays: Design Issues for Rotary-Wing Aircraft* (Rash, 2001). The following discussion steps back from some of the details of these systems and looks at a bigger picture to identify themes that will apply to many different situations.

Although engineering teams do not make an HMD awkward to use, HMDs often fail to live up to their promised performance (Keller and Colucci, 1998). Many of the issues with HMDs are those common to all information displays, which can either make information more useable or can increase workload and stress (Gilger, 2006). In spite of conscientious efforts by the human factors engineering (HFE) community, HMD designs have not been optimized for the capabilities and limitations of the human user (National Research Council, 1997). The progress that has been made in addressing HFE issues has been modest and largely limited to either anthropometry or to the physiological characteristics of the human senses, i.e., vision and audition. Perceptual and cognitive factors associated with HMD user performance have been almost totally overlooked. With the information-intensive modern battlespace, these factors are taking on an even greater importance.

While HMDs can be used for a wide variety of purposes and display many different types of information, fundamentally there is always a "region" where a human user interacts with the HMD. This is the *human-machine interface* (HMI). This interface serves as a *bridge* that connects the user and the machine (Hackos and Redish, 1998). The design of this interface is critically important because the information from quality sensors and computer analysis will not be beneficial unless the human user understands the information. It is important to note that the HMI is not a device; instead, it is a virtual concept, represented by the interaction of the human sensory, perceptual and cognitive functions with the HMD's information output(s).

This chapter is organized to examine the different aspects of the human-machine interface. We start with a basic description of human perceptual and cognitive systems, and consider their biases, abilities, and limitations. We then turn to a description of HMDs and consider their abilities and constraints. Finally, we discuss the interface between these two systems and consider general aspects of how they can be brought together. This discussion is kept at a relatively high-level abstraction of ideas and leaves the details for other chapters.

Human Sensation, Perception and Cognition

Sensation, perception, and cognition all refer to the acquisition, representation, and utilization of information in the world. These processes appear easy, automatic, and complete, but in reality, they are extremely complex, take substantial processing, and are surprisingly limited in terms of their relation to the veridical world.

Sensation

Sensation is one of the first steps in acquiring information about an environment. It refers to the detection of a property (or characteristic) of an object in the world. Typically this process involves responses from biological receptors that are sensitive to a particular form of energy. These receptors can be very complex and can respond to a wide range of energy forms. For vision, the receptors are cells in the back of the eye called rods and cones that respond to light energy of different wavelengths. For audition, the receptors are the cilia of the organ of Corti that sit on the basilar membrane in the cochlea of the ear. For cutaneous sensation (touch), there are several types of receptors that are embedded in the skin and respond to flutter, vibration, pressure, and stretching. Figure 2-1 shows schematic views of the receptors for vision, audition, and touch.

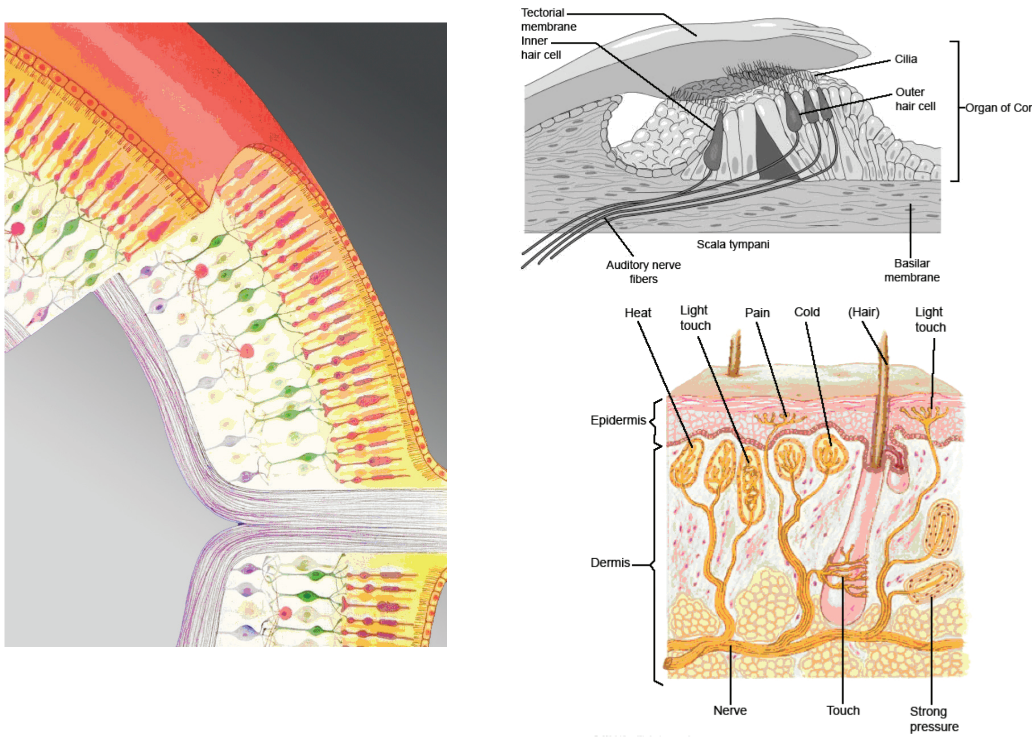

Figure 2-1. Different receptors are responsible for the sensation of dissimilar types of stimulus energy. Left: Cross section of the back of the eye shows photoreceptors that are sensitive to light energy. Top right: Cilia on the organ of Corti are sensitive to sound energy. Bottom right: Receptors in the skin are sensitive to forces on the skin.

Perception

Humans are not aware of sensory processes, except as they influence our perception of the world. Perception refers to the awareness of objects and their qualities. The process of perception is so accurate and convincing that the detailed mechanisms of how perception happens are mostly hidden from general awareness. We have the impression that as soon as we open our eyes, we see the world, with all of its objects, colors, patterns, and possibilities. In reality, the events that occur when the eyes open are astonishingly complex processes that depend on precise chemical changes in the eye, transmission of electrical and chemical signals through dense neural circuits, and rich interactions with both memories of previous events and planned interactions with the world.

Figure 2-2 characterizes the *perceptual loop* that is mostly hidden from awareness when looking at the world. This figure and the following discussion are adapted from Goldstein (2007). (See Chapter 15, *Cognitive Factors*, for a similar loop that describes some cognitive processes.) One could start a description of the loop at any place and could talk about any of the perceptual senses. We will focus on visual perception because it is easy to refer to the stimuli, and we will start with the *Action* node on the far right. Here, a human interacts with the environment in some way that changes the visual array. This could be as simple as opening the eyes, turning the head, or taking a step forward. It could also be a quite complex event such as jumping toward a ball, splashing paint on a surface, or changing clothes. The action itself changes the environment. Thus, the next step in the loop is the *Environmental stimulus*.

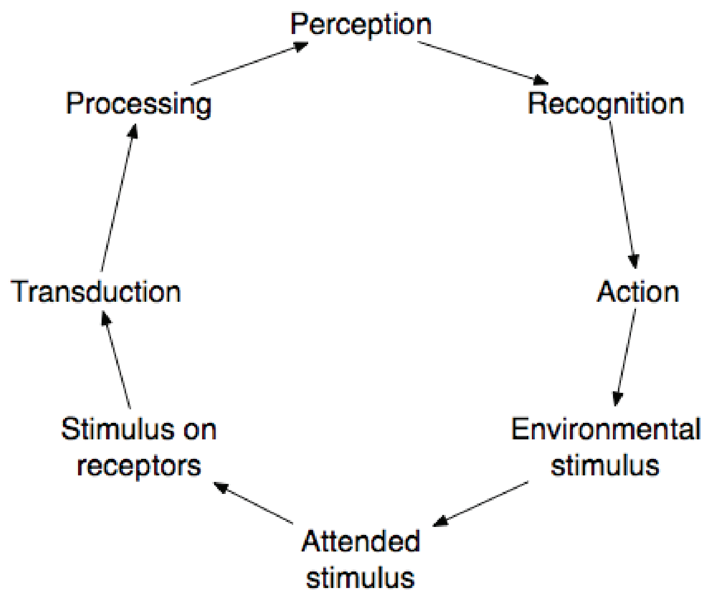

Figure 2-2. The perceptual loop demonstrates that perceptual processing involves many different complex interactions between the observer and the environment (adapted from Goldstein, 2007).

The environmental stimulus refers to properties of things in the world. This is the information that is, in principle, available to be acquired. For visual perception, this refers to the currently visible world. In practice, a person cannot acquire information about the entire environmental stimulus. Instead, perceptual processes usually focus on only a relatively small subset of the environmental stimulus, the *attended stimulus*.

The attended stimulus is the part of the environmental stimulus that is positioned in such a way that sensory systems can acquire information. The term *stimulus* may need some explanation, as it is very context specific. In some situations, the stimulus may be a particular object in the world, such as a building. In other situations the stimulus may refer to a particular feature of an object in the world, such as the color of a building's wall. In still other situations, the stimulus may be a pattern of elements in the world, such as the general velocity and direction of a group of aircraft. A person can attend a stimulus by moving the body, head, and eyes so that the relevant parts of the environmental stimulus fall on to the appropriate perceptual sensors.

The *stimulus on receptors* is the next step in the perceptual loop. Here, energy that corresponds to the attended stimulus (and some energy from parts of the environmental stimulus that are not the attended stimulus) reaches specialized cells that are sensitive to this energy. For visual perception, the specialized cells are photoreceptors in the back of the eye (Figure 2-1). These photoreceptors are sensitive to light energy (photons).

Transduction refers to the conversion of stimulus energy into a form of energy that can be used by the nervous system of the observer. The stimulus energy must be converted into an internal signal that encodes information about the properties of the attended stimulus. For visual perception, the photoreceptors of the eye initially undergo chemical transformations when they absorb photons. These chemical transformations induce an electrical change in the photoreceptor. These electrical changes are then detected and converted into a common format that the rest of the brain can use.

After transduction, the stimulus undergoes *processing* by the neural circuits in the brain. This processing includes identification of stimulus features (e.g., patterns of bright and dark light; and oriented edges of stimuli). As this information is processed, the observer experiences the phenomenological experience of *perception*. At this stage, the observer gains an awareness of properties of the attended stimulus in the world. The perceptual experience is not simply a function of the attended stimulus energy, because the experience may also depend on a memory of the world from a few previous moments. It may also depend on the action generated by the observer.

Recognition refers to additional interactions with memory systems to identify the attended stimulus, relative to the observer's experience and current needs. Here, the observer interprets the properties of the attended stimulus, perhaps to identify friend or foe and opportunity or threat. As a result of this interpretation, the observer generates some kind of action, which restarts the perceptual loop.

While we have stepped through the stages of the perceptual loop one after another, in reality all the stages are operating simultaneously and continuously. Thus, actions based on one moment of recognition may occur at the same time as transduction from a previous stimulus on receptors. Moreover, some information about the environment can only be detected after multiple passes through the perceptual loop, where the observer plans a specific sequence of actions so that they change the environmental stimulus in a way that allows them to gain particular desired information (e.g., moving the head back and forth to induce a motion parallax, which allows for discrimination of an object in depth).

One of the main messages from the description of the perceptual loop is that perception is an extremely complex experience. Each stage of the perceptual loop plays an integral role in perceptual experience and contributes to how we interpret and interact with the world. What is known about the details of each stage in the perceptual loop is far too complicated to describe in this book. Some of the other chapters in this book do discuss some of the details that are especially important for HMDs. Here, we try to take a more global view of the issues.

The human perceptual systems have evolved to process only certain types of stimulus inputs. For example, the human visual system covers only a small subset of the electromagnetic spectrum (i.e., 380 to 730 nanometers). We interpret different wavelengths of light as perceptually different colors, but the visual system is unaware of electromagnetic energy at longer wavelengths (heat) or very short wavelengths (ultraviolet and beyond).

Similarly, the human visual system has evolved to detect subtle properties of the visual world by interpreting global flows of streaming motion (Gibson, 1950). As we move through an environment, individual objects in the world produce moving patterns of light across our retinas. The patterns of movement contain significant information about the world and the properties of the observer. Figure 2-3 schematizes two flow fields generated by different movement of the observer. The line projecting out from each dot indicates the direction and velocity (length of the line) of a dot at that position in the field-of-view (FOV). Figure 2-3A shows the flow field generated when the observer moves in a straight line toward a fixed point in the middle of the field. All of the motion patterns expand from the fixed point. Sensitivity to the properties of the flow field can allow a moving observer to be sure that he or she is moving directly toward a target.

Flow fields can be much more complicated. Figure 2-3B shows the flow field generated by an observer traversing on a curved path while fixating on the same spot as in Figure 2-3A. To maintain fixation on a point, the observer must change his or her head or eyes, and these movements change the properties of the flow field.

Humans can use these kinds of flow fields to estimate heading direction to an accuracy within one visual degree (Warren, 1998), and many areas of the brain are known to be involved in detecting motion and flow fields (Britten and van Wezel, 1998). Flow fields of this type exist for many different situations, and they are especially important for detecting heading and direction of motion in aircraft (Gibson, Olum and Rosbenblatt, 1955). However, there are some kinds of flow fields that humans interpret incorrectly and so produce perceptual illusions (e.g., Fermuller, Pless and Aloimonos, 1997). Thus, the perceptual systems limit the kinds of information that people can extract from flow fields.

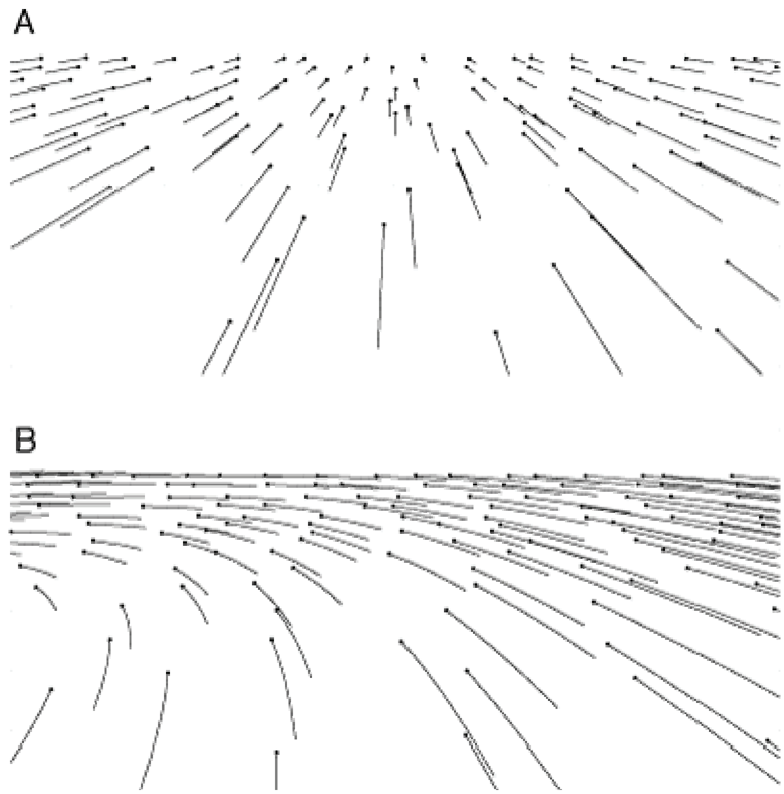

Figure 2-3**.** A) Radial dot flow generated from a straight-line path across a ground plane. The direction of motion can be determined by finding the focus of expansion, the point in the flow field where there is no horizontal or vertical motion. This may not be explicitly present, but can be extrapolated from the motion of other points in the image. B) Curvilinear dot flow generated from a curved path across a ground plane, also with a fixed gaze. From Wilkie and Wann (2003).

There are similar issues for depth perception. Objects in the world occupy three spatial dimensions, but the pattern of light on the retina of an eye is a 2-dimensional projection of light from the world. The third dimension must be computed from the differences in the projection to the two eyes, by changes in the projection over time (motion parallax), or by pictorial cues that generally correlate with differences in depth. As part of this process, the human visual system has evolved to make certain assumptions about the world. These assumptions bias the visual system to interpret properties of a scene as cues to depth. For example, the objects in the top row of Figure 2-4 generally look like shallow holes, while the objects in the bottom row look like small hills. There is a bias for the visual system to assume that light sources come from above objects. The interpretation of the objects as holes is consistent with this idea. Now rotate the page so that the figure is upside down. The same bias for light to come from above now switches the percept of the items.

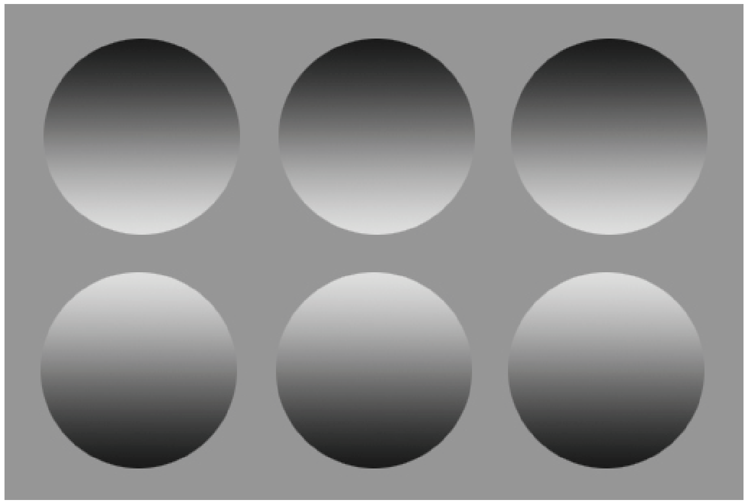

Figure 2-4. The visual system is biased to assume an illuminant comes from above. The perceived depths associated with the dots reveal this bias. The top row appears to be made of shallow holes because the brighter part is at the bottom and the top is darker (in shadow). The bottom row appears to be made of small hills because the top is brighter (light hitting it) while the bottom is darker (in shadow).

There are similar biases for interpreting patterns of reflected light. Figure 2-5 shows what is called the Snake Illusion (Adelson, 2000; Logvinenko et al., 2005). The diamonds are all the same shade of gray, but are set against light or dark backgrounds. They look different because the visual system interprets the dark bar on top as a transparent film in front of the gray diamonds and the white background. As seen through such a film, the gray diamonds appear brighter than the (physically identical) gray diamonds below that are not seen through a film. Here, a bias in the visual system to interpret patterns of light as indicative of transparent surfaces changes the apparent brightness of objects. Such complex interpretations of scenes are quite common (Gilchrist et al., 1999).

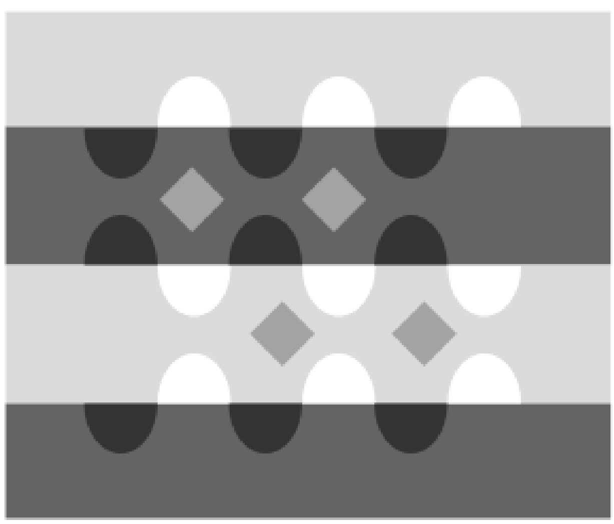

Figure 2-5. The Snake Illusion: The gray diamonds are physically the same shade of gray, but the diamonds in the top row appear lighter than the diamonds in the bottom row. Adapted from Adelson (2000).

Although more difficult to demonstrate in a printed format, there are similar biases and influences for other perceptual systems. Humans detect sounds only within a certain band of frequencies, and have varying sensitivities across these frequencies that are biased toward the range of frequencies that correspond to human speech (Fletcher and Munson, 1933). Likewise, segmentation of the auditory stream follows certain rules that can cause a listener to perceive multiple sound sources when only one sound source is actually present. See Chapters 11, *Auditory Perception and Cognitive Performance*, and 13, *Auditory Conflicts and Illusions*, for further discussions of human auditory perception, and Chapter 18, *Exploring the Tactile Modality for HMDs*, for a description of haptic perception.

The main lesson from these observations about human perception is that the perceptual systems have evolved to identify and extract some types of information in the environment but have not evolved to process other types of information. Evolutionary pressures have lead to perceptual systems that operate well within some environments, but these same systems will behave poorly when placed in entirely new environments.

Cognition

Similar observations can be made about human cognition (see Chapter 15, *Cognitive Factors*, for a fuller discussion of cognitive systems). Humans are very good at tasks involving face recognition (e.g., Walton, Bower, and Power, 1992) because evolutionary pressures give an advantage to being able to recognize, interpret, and remember faces. Humans also are quite good at many pattern recognition tasks that are difficult for computers, such as reading handwriting, interpreting scenes, or understanding speech in a noisy environment (Cherry, 1953). However, there are many recognition tasks where humans perform quite poorly, especially tasks that involve judgments of probability or the use of logic (Khaneman and Tversky, 1984). Moreover, biases and limitations of perceptual, attentional, memory, decision-making, and problem solving systems severely restrict the ability of individuals to perform well in many complex situations.

We complete this description of human behavior by pointing out a few common misconceptions about perception and cognition. First, evolutionary pressures rarely lead to optimal behaviors, and humans rarely act in an optimal way. Instead, evolution tends to select solutions that satisfy many different constraints well enough. Humans are good pattern recognizers, but outside of a few special situations it would be false to characterize them as optimal. Second, perception does not involve direct awareness of the world. Some researchers go so far as to claim that all of perception is an illusion, but this presupposes that one has a good definition of reality. Such philosophical discussions (Sibley, 1964) are beyond the scope of this book, so we simply note that perception actually requires significant resources and processing to acquire information about an environment. Third, contrary to centuries of philosophizing, humans are not generally rational. Studies of human cognition show that when humans appear to be rational it is not because they think logically, but because they learn the specific rules of a specific situation and act accordingly (Wason and Shapiro, 1971). Thinking rationally requires substantial training in the rules of logic, and this often does not come naturally. Finally, it is a mistake to believe that an individual can use all available information. The presence of information on a display or “known” to a person does not mean that such information or knowledge will guide human behavior in any particular situation.

Machine: Helmet-Mounted Displays

HMDs can be constructed in many different ways. Variations in sensors can make an image on a display sensitive to different aspects of the environment. Variations in the display change how information is presented to the human observer. Whatever the application, HMDs are not stand-alone devices. As integrated components of combat systems (as well as in other applications), they are used to present information that originates from optical and acoustic sensors, satellites, data feeds, and other communication sources, Even in simulation or virtual immersion applications, external signals (consisting of visual and audio data or information) must be provided. In the following discussion, we briefly place the HMD in perspective, by considering its role as just one component of the night imaging system. The function of the HMD does not come to bear until energy that is created by or reflected from objects and their environment (referred to as *stimuli*) is captured (detected) by *sensor(s)*, and then manipulated, transmitted, and presented on the HMDs *displays*. While not an exhaustive examination of the important properties of HMDs, this distinction helps to highlight some of the key features that relate to the HMI.

Stimuli

There are several ways to define a stimulus, but usually the term is used to refer to the properties of objects of interest in the world that generate sensations. This definition is important because an HMD filters and modifies the detected properties of the object. Thus, for example, a faint visual stimulus that normally would be undetected by the unaided eye can become detected with the aid of a night vision sensor; similarly, a faint sound stimulus that normally would be undetected by the unaided ear can become detected with the aid of an amplifier. A different way to describe the situation is to note that the night vision sensor converts one stimulus (the original faint stimulus) into another stimulus (visual energy in the HMD's display component). These are largely philosophical distinctions, although it is sometimes useful to switch between descriptions to explain different aspects of perception.

For human vision, input sources can be any object that emits or reflects light energy anywhere in the electromagnetic spectrum. For nighttime operations, examples include obvious naked-eye sources such as weapon flashes, explosions, fires, etc., and thermal sources such as human bodies, tanks, aircraft, and other vehicles that would serve as emissive sources during and after operation.

For human hearing, input sources are both outside and inside the personal space (e.g., cockpits for aviators and vehicle interiors for mounted Warfighters). Outside audio input sources include explosions, weapon fire, and environment surround sounds (especially for dismounted Warfighters). Inside sources include engine sounds, warning tones, and communications.

With an HMD application, properties of the external world are detected by sensors and are then converted into electronic signals. These signals are relayed to the visual or audio display component of the HMD, where an image of the external "scene" (visually or acoustically) is reproduced, sensed, and then acted upon by the user. A simplified block diagram for this visual/acoustical stimulus-sensor-display-user construct is presented in Figure 2-6. In this simplistic representation, the HMD acts as a platform for mounting the display (or, in some designs, a platform for mounting an integrated sensor/display combination).

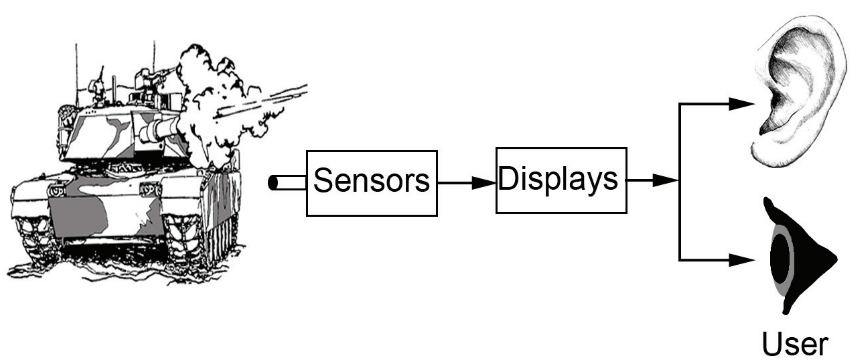

Figure 2-6**.** Simplified block diagram of the visual/acoustical stimulus-sensors-displays-user construct used in HMDs.

Sensors

Sensors are devices that acquire information about stimuli in the outside world. A sensor is defined as a device that responds to a stimulus, such as heat, light, or sound, and generates a signal that can be measured or interpreted. HMD visual imagery is based on optical sensors. Historically, the use of acoustic sensors in the battlespace has been limited, with underwater applications being the most prevalent. Generally, HMD audio information presentation has been limited to the reproduction of communications via speakers. However, acoustic sensors are rising in importance as their utility is explored. Different acoustic sensors operating in the ultrasonic and audible frequency ranges have a wide range of applications and impressive operating ranges. Optical forward-looking infrared (FLIR) imaging sensors can have an effective detection operating range as great as 20 kilometers (km) (12.4 miles) under optimal environmental conditions; acoustic sensors theoretically can operate out to approximately 17 km (10.6 miles) under ideal conditions. For both sensor technologies, identification ranges are more limited.

Many HMDs are based on optical imaging systems and are used to augment normal human vision. These systems include sensors that are sensitive to energy that is not detected by the normal human eye. The HMD displays this energy in a way that helps an observer identify objects in the environment. Optical imaging sensors can be categorized by the type of energy (i.e., range of wavelengths) they are designed to detect. Each specific category defines the imaging technology type (and therefore the physics) used to convert the scene of the external world into an image to be presented on the HMD's display. Theoretically, such sensors may operate within any region of the electromagnetic spectrum, e.g., ultraviolet, visible, IR, microwave, and radar. Currently, the two dominant imaging technologies are image intensification (I^2) and FLIR.

Image intensification (I^2) sensors

The sensor used in an I^2 system (as applied in early generation I^2 devices) uses a photosensitive material, known as a photocathode, which emits electrons proportional to the amount of light striking it from each point in the scene. The emitted electrons are accelerated from the photocathode toward a phosphor screen by an electric field. The light emerging from the phosphor screen is proportional to the number and velocity of the electrons striking it at each point. The user views the intensified image formed on the phosphor screen through an eyepiece (Figure 2-7).

I^2 sensors generally detect energy in both the visible range and the near-IR range; the actual wavelength range is dependent on the technology generation of the I^2 sensor (and sometimes the presence of optical filters).

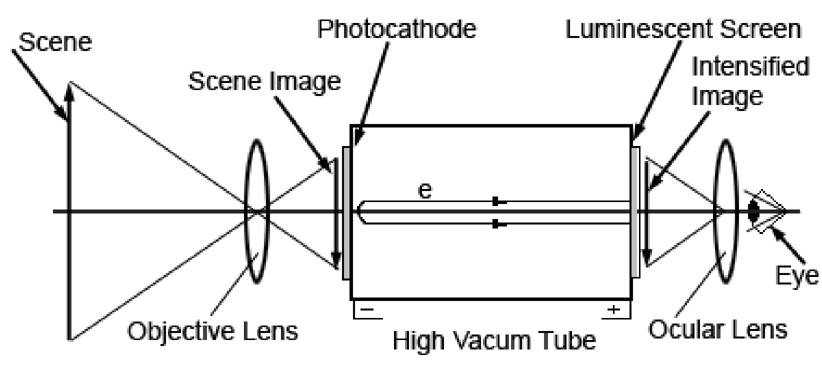

Figure 2-7. The basic parts of an I^2 device. The photocathode effectively amplifies the light intensity of the visual scene and projects the amplified scene on a screen that is perceived by the observer.

This process is analogous to using a microphone, amplifier and speaker to allow the user to more easily hear a faint sound. In both cases, some of the “natural fidelity” may be lost in the application process. The intensified image resembles a black-and-white television image, only usually in shades of green (based on the selected display phosphor) instead of shades of gray. However, recent advances offer the promise of pseudo-color I^2 devices based on dual- or multi-spectrum I^2 technology (Bai et al., 2001; Toet, 2003; Walkenstein, 1999).

Forward-looking infrared (FLIR) sensors

FLIR-based imaging systems operate on the principle that every object emits energy (according to the Stefan-Boltzmann Law). The emitted energy is a result of molecular vibration, and an object’s temperature is a measure of its vibration energy. Therefore, an object’s energy emission increases with its temperature. An object’s temperature depends on several factors: its recent thermal history, its reflectance and absorption characteristics, and the ambient (surrounding) temperature.

FLIR sensors detect the IR emission of objects in the scene and can “see” through haze and smoke and even in complete darkness. Although no universal definition exists for infrared (IR) energy, for imaging purposes, it is generally accepted as thermally emitted radiation in the 1 to 20 micron region of the electromagnetic spectrum. Currently, most military thermal imaging is performed in the 3 to 5 or 8 to 12 micron region. These regions are somewhat dictated by the IR transmittance windows of the atmosphere (Rash and Verona, 1992).

Thermal imaging sensors form their image of the outside world by collecting energy from multiple segments of the outside scene. The sensors convert these energy data into a corresponding map of temperatures across the scene. This may be accomplished by using one of several sensor designs. The two most common are the older *scanning arrays* and the newer *focal plane staring arrays*.

Typically a scanning array consists of a vertical row of sensor elements. This 1-D array is scanned horizontally across the focused scene, producing a 2-D profile signal of the scene. If desired, the scan can be reversed to provide an interlaced signal.

A focal plane array uses a group of sensor elements organized into a rectangular grid. The scene is focused onto the array. Each sensor element then provides an output dependent upon the incident infrared energy. Temperature resolution, the ability to measure small temperature differences, can be as fine as 0.1° C.

Acoustic sensors

The source inputs for auditory displays are often thought of as being radio-transmitted communications, automated voice commands, or artificially generated alert tones. Historically, any sensing of external sounds, such as engine sounds, weapons fire, ground vibration, etc., has been accomplished primarily by the human ear, and for lower frequencies, the skin. However, in the last decade there has been an increased interest in acquiring external sounds and using them for source identification and spatial localization, e.g., sniper fire from a specific angle orientation. This is accomplished through the use of acoustic sensors.

Acoustic sensor technology involves the use of microphones or arrays of microphones to detect, locate, track, and identify air and ground targets at tactical ranges. Target information from multiple widely-spaced acoustic sensor arrays can be digitally sent to a remote central location for real-time battlespace monitoring. In addition, acoustic sensors can be used to augment the soldier's long range hearing and to detect sniper and artillery fire (Army Materiel Command, 1997).

Acoustic sensors have been used for decades in submarines for locating other submarines. The earliest and most familiar is the “hydrophone.” The hydrophone is a device that detects acoustical energy underwater, similar to how a microphone works in air. It converts acoustical energy into electrical energy. Hydrophones in underwater detection systems are passive sensors, used only to listen. The first hydrophone used ultrasonic waves. The ultrasonic waves were produced by a mosaic of thin quartz crystals placed between two steel plates, having a resonant frequency of approximately 150 kilohertz (kHz). Contemporary hydrophones generally use a piezoelectric ceramic material, providing a higher sensitivity than quartz. Hydrophones are an important part of the Navy SONAR (SOund Navigation And Ranging) systems used to detect submarines and navigational obstacles. Directional hydrophones have spatial-localized sensitivity allowing detection along a specific direction. Modern SONAR has both passive and active modes. In active systems, sound waves are emitted in pulses; the time it takes these pulses to travel through the water, reflect off of an object, and return to the ship is measured and used to calculate the object's distance and often its surface characteristics.

No longer confined to the oceans, modern acoustic sensors are being deployed by the U.S. Army in both air and ground battlespace applications. These sensor systems have demonstrated the capability to detect, classify, and identify ground targets at ranges in excess of 1 km and helicopters beyond 5 km with meter-sized sensor arrays, while netted arrays of sensors have been used to track and locate battalion-sized armor movements over tens of square kilometers in non-line-of-sight conditions.

Regardless of the characteristics of individual sensor types, the sensors employed by an HMD system are designed to detect certain types of energy and present information about that energy to the user. The properties of the system are thus fundamentally defined by the sensitivity of the sensors to the energy they detect. For an observer to respond to this energy, the sensors must convert the detected energy into a format that can be detected by the observer's perceptual system. The converted energy is then displayed to the observer. This leads to the remaining main component of any HMD-the display.

Displays

A generic definition of a "display" might be "something used to communicate a particular piece of information." A liberal interpretation of this definition obviously should be extremely broad in scope. Examples would run the gamut from commonplace static displays (e.g., directional road signs, advertising signs, posters, and photographs) to dynamic displays (e.g., televisions, laptop computer screens, and cell phone screens).

Visual displays used in HMDs are so ubiquitous that they almost do not need any introduction. There are many different type of visual displays (e.g., cathode-ray tubes, liquid crystal, electroluminescent, etc.), but they all generate patterns of light on a 2-dimenional surface. Details about the properties, constraints, and capabilities of these display types are provided in Chapter 4, *Visual Helmet-Mounted Displays*.

Because they are less familiar to many people, we will describe auditory displays in more detail. Auditory displays use sounds to present information. These sounds can be speech-based, as with communications systems, or nonspeech-based, such as the "beep-beep" of a microwave oven emitted on completion of its heating cycle. Auditory displays are more common than might first be assumed. They are used in many work environments including kitchen appliances, computers, medical workstations, automobiles, aircraft cockpits, and nuclear power plants.

Auditory displays that use sound to present data, monitor systems, and provide enhanced user interfaces for computers and virtual reality systems are becoming more common (International Community for Auditory displays, 2006). Examples of auditory displays include a wide array of speakers and headphones.

Auditory displays are frequently used for alerting, warnings, and alarm-situations in which the information occurs randomly and requires immediate attention. The near omni-directional character of auditory displays that can be provided using an HMD is a major advantage over other types of auditory displays.

Long used primarily as simple alerts, the presentation of nonspeech-based sounds is increasing in its scope, effectiveness and importance. Sound is being explored as an alternate channel for applications where the presence of vast amounts of visual information is resulting in "tunnel vision" (Tannen, 1998). However, sound is sufficient in its own capacity to present information.

There is a vast spectrum of sounds available for use in auditory displays. Kramer (1994) describes a continuum of sounds ranging from audification to sonification. Audification refers to the use of "earcons" (Houtsma, 2004), a take-off on the concept of icons used in visual displays. An icon uses an image that "looks" like the concept being presented, e.g., a smiley face representing happiness; an earcon would use a sound that parallels that of the represented event. These are typically familiar, short-duration copies of real-world acoustical events. As an example, an earcon consisting of a sucking sound might be used in a cockpit to warn the aviator of fuel exhaustion.

Sonification refers to the use of sound as a representation of data. Common examples of sonification use include SONAR pings to convey distance information about objects in the environment and the clicks of a Geiger counter to indicate the presence of radioactivity. In both examples the sounds are the means of information presentation; but the actual sounds themselves are not meaningful. Instead, it is the relationship between the sounds that provide information to the user (in the case of SONAR, the relationship is between distance and time; for the Geiger counter, the relationship is between intensity and frequency).

It is worth reemphasizing that audification uses the structure of the sound containing the information, while sonification uses the relationship of sounds to convey the information. This implies that in the case of sonification, changing the specific sounds does not change the information, and even simple tones may be employed (Kramer, 1994).

The development of 3-D auditory displays for use in HMDs is both an example of the sophisticated level that auditory displays have achieved and an example of an application where an auditory display is superior to a visual display. The inherent sound localization produced by such displays can be used to directionally locate other Warfighters (friend and foe), threats, and targets (Glumm et al., 2005).

Auditory display technologies for HMD applications are not as diverse as visual display technologies. The dominant technology is the electro-mechanical or electro-acoustic transducer, more commonly known as a speaker, which converts electrical signals into mechanical and then acoustical (sound) signals. More precisely, it converts electrical energy (a signal from an amplifier) into mechanical energy (the motion of a speaker cone). The speaker cones, in turn, produce equivalent air vibrations in order to make audible sound via sympathetic vibrations of the eardrums.

An alternate method of getting sound to the inner ear is based on the principle of bone conduction. Headsets operating on this principle (referred to also as ears-free headsets) conduct sound through the bones of the skull (cranial bones). Such headsets have obvious applications for hearing-impaired individuals but have also been employed for normal-hearing individuals in auditory-demanding environments (e.g., while scuba diving) (MacDonald et al., 2006).

Bone conduction headsets are touted as more comfortable, providing greater stereo perception, and being compatible with hearing protection devices (Walker and Stanley, 2005). However, bone conduction acts as a low-pass filter, attenuating higher frequency sounds more than lower frequency sounds.

Auditory displays, their technologies and applications, are discussed further in Chapter 5, *Audio Helmet-Mounted Displays*.

Other issues

HMD systems face additional constraints because they are almost always a part of a larger system. In military settings, HMDs are almost always a part of the Warfighter's head protection system (i.e., helmet). As a result, the HMD must not introduce components that undermine the head protection crash worthiness of the system, e.g., impact and penetration protection (see Chapter 17, *Guidelines for HMD Designs*). One effect of this constraint is that the HMD components face restrictions on their weight and how their placement affects the center-of-mass of the combined HMD/helmet system.

Another issue that drives an HMD system design is how it interacts with the environment in which it is to be used. A key aspect is that the HMD needs to be relatively self-contained. That is, the HMD must be able to operate with a system that may change in several significant ways. While one wants the HMD to match appropriately with the larger system, it is not practical for a minor change in the larger system to necessitate a major redesign of the HMD. In addition to working well with various types of machine systems, an HMD needs to work well with various types of human users. While people's cognitive and perceptual systems are fairly similar, there can be significant differences, and the HMD needs to be functional for a variety of users. User capabilities may change over time, and an HMD needs to be usable despite these changes. Even with the best HMD development programs, system and user performance are usually evaluated under generally benign conditions and not under more realistic conditions where physical fatigue, psychological stress, extreme heat or cold, reduced oxygen levels, and disrupted circadian rhythms are present.

The HMD as a Human-Machine Interface: Statement of the Challenge

The task of designing an HMD that both meets the needs of the situation and matches the abilities of the user is a difficult one. The best efforts and intentions may still lead to a poor result. There are so many constraints on an HMD from the physical, task, and human parameters that something is almost certain to be suboptimal. Unfortunately, for the system to work well, everything must be just right.

Many of the difficulties derive from the need for an HMD to behave robustly in a complex system. From engineering and manufacturing perspectives, an HMD needs to be relatively self-contained. Unless the HMD behaves robustly, the manufacture or design of the system components can bog down development. For example, changes to one part of an HMD system (e.g., a microphone) must be relatively isolated from other parts of the HMD (e.g., the visual display).

Having described the human and machine aspects of an HMD, we are now ready to discuss how the properties of these two systems influence the design of the human machine interface. The HMI challenge is to address the following question: *How to use robust technology to organize and present information in a way that meets the expectations and abilities of the user?*

Clearly, a satisfactory solution to the challenge requires careful consideration of both the machine and human systems. Current engineering techniques tend to focus on ensuring that the machine side of the system behaves according to design specifications in a way that ensures that appropriate sensor information is present on the display. There are remaining issues to be resolved, and active development of new technologies will be needed to address these issues. For example, a continued effort in the development of miniature display technologies can improve weight, center-of-mass offset and heat generation, which in turn improves comfort. Development of more intuitive symbology (an ongoing effort) will reduce workload and error rate.

The more difficult aspect of the challenge, and the part that needs more progress, is understanding the human side of the system. Information on an HMD may be present but not be perceived, interpreted, or analyzed in a way that allows the human user to take full advantage of the HMD. Working with the human side is difficult because many aspects of human perception and cognition are not fully understood and thus there is little information available to guide the design of an HMD. Moreover, humans are exceptionally complex systems that can behave in fundamentally different ways in different contexts. These behavior changes make it very difficult to predict how they will perform in new situations. Indeed, one commonly noted aspect of fielded HMD designs is that users do not follow the "rules" for the system and instead adapt new strategies to make the HMD operate in some unexpected way. A classic example of HMD users not following the rules is AH-64 Apache pilots using the Integrated Helmet and Display Sighting System (IHADSS). This HMD has a very small exit pupil that results in great difficulty maintaining the full FOV. To compensate, pilots use a small screwdriver to minimize the image on the display, thereby allowing viewing of the full FOV (but no longer in a one-to-one relationship with the sensor FOV) (Rash, 2008).

There has been substantial progress on some aspects of the challenge. For example, studies of human vision indicate the required luminance levels that are needed for HMD symbology to be visible in a wide variety of background scenes. Likewise, the intensities and frequencies of sound stimuli that can be detected by human users are well understood and promote guidelines for HMD design (Harding et al., 2007).

Things are more challenging when the information detected by sensors does not correspond to aspects of the world that are usually processed by human perceptual systems. For example, infrared vision systems that detect sources of heat energy can provide a type of "night vision." The information from sources of heat energy is usually displayed as a visual display. Such a display requires some type of conversion from heat energy to the visible ranges of light energy. This conversion can lead to misinterpretations of the information when aspects of the sensor information are mapped onto display properties in a way that is inconsistent with the biases of the visual or cognitive systems. In the case of heat sensors, it is fairly easy to display the intensity of heat emissions as a light intensity map. This provides the human observer with an unambiguous description of what the sensor has detected. However, a light intensity map tends to be interpreted as something produced by objects that reflect illuminated light. As a result, the visual display can be misinterpreted, with columns of heat interpreted as solid objects and false identification of figure and ground.

Adding color to the display of such a system may provide additional clarity about the properties of the heat emissions, but can lead to even further confusion about the properties of the objects in the environment. In normal vision, different colors correspond to changes in the properties of surfaces (e.g., fruit versus leaves), but may correspond to something else entirely on a visual display.

Thus, the great benefit of HMDs, that they can display a wide array of sensor information, also exposes them to great risk, that they display information in a way that is inconsistent with the properties of the observer.

In optimizing the HMI for the HMD, the electrical engineer might investigate how to build better buttons and connectors (or other physical components of the HMD); the human factors engineer might investigate how to design more legible/audible and intelligible labels, alerts or instructions (e.g., perhaps, the characteristics of the symbology presented via the HMD); the ergonomicist might investigate the anatomy and anthropometry of the user population (e.g., head dimensions and interpupillary distance); but in this book we will focus on investigating HMD design from the perspective of the *in toto* human visual, auditory and neural systems (i.e., sensory, perceptual and cognitive functions). In doing so, the bidirectional flow of information will be studied via the HMD, through the sense organs (primarily the eyes and ears), through the visual and auditory pathways, through the thalamus, to and from the respective cortices. The HMI concept adopted here will incorporate the relationship between the HMD design and the user's visual and auditory anatomy and physiology, as well as the processes by which we understand sensory information (perception) and the neural activities associated with recognition, memory, and decision making with this information (cognition).

All of the issues are addressed in the following chapters. There is, as yet, no complete solution to the HMI challenge, but progress is being made in many areas. One goal of this book is to identify where solutions do exist, identify situations that require additional study, and outline possible solutions to some of those problem situations.

References

Adelson, E.H. (2000). Lightness perception and lightness illusions. In Gazzaniga, M. (Editor), *The New Cognitive Neurosciences*. Cambridge, MA: MIT Press, 339–351.

Army Materiel Command. (1997). *Command and Control Research Program. Army Science and Technology Master Plan (ASTMP).* Alexandria, VA: Army Materiel Command.

Bai, L., Gu, G., Chen, Q., and Zhang, B. (2001). Information obtaining and fusion of color night vision system. *Data Mining and Applications, Proceedings of SPIE*, 4556, 65-70.

Britten, K.H., and Wezel, R.J.A. van. (1998). Electrical microstimulation of cortical area MST biases heading perception in monkeys. *Nature Neuroscience*, 1, 59-63.

Cherry, E.C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America*, 25(5), 975-979.

Fermüller, C., Pless, R., and Aloimonos, Y. (1997). Families of stationary patterns producing illusory movement: Insights into the visual system, *Proc. Royal Society London B*, 264, 795-806.

Fletcher, H., and Munson, W.A. (1933). Loudness: Its definition, measurement, and calculation. *Journal of the Acoustical Society of America*, 5, 82-108.

Gibson, J.J. (1950). *Perception of the Visual World.* Boston: Houghton Mifflin.

Gibson, J.J., Olum, P., and Rosenblatt, F. (1955). Parallax and perspective during aircraft landings, *American Journal of Psych*ology, 68, 372-385.

Gilchrist, A., Kossyfidis, C., Agostini, T., Li, X., Bonato, F., Cataliotti, J., Spehar, B., Annan, V., and Economou, E. (1999). An anchoring theory of lightness perception. *Psychological Review*, 106(4), 795-834.

Gilger, M. (2006). Information display-the weak link for NCW. *Defense, Security, Cockpit, and Future Displays II, Proceedings of SPIE*, 6225, 622511-11 to 622511-12.

Glumm, M.M., Kehring, K.L., and White, T.L. (2005). Effects of tactile, visual, and auditory cues about threat location on target acquisition and attention to visual and auditory communications. *Human Factors and Ergonomics Society Annual Meeting Proceedings, Cognitive Engineering and Decision Making,* 347-351.

Goldstein, E.B. (2007). *Sensation and Perception, 7th Edition*. Pacific Grove, CA : Thomson Wadsworth.

Hackos, J., and Redish, J. (1998). *User and task analysis for interface design*. New York: John Wiley and Sons.

Harding, T.H., Martin, J.S., and Rash, C.E. (2007). The legibility of HMD symbology as a function of background local contrast. *Head- and Helmet-Mounted Displays XII: Design and Applications, Proceedings of SPIE*, 6557, 622570D.

Houtsma, A. (2004). Nonspeech audio in helicopter aviation. Fort Rucker, AL: U.S. Army Aeromedical Research Laboratory. USAARL 2004-03.

International Community for Auditory Displays (ICAD). (2006). What is ICAD? Retrieved May 10, 2007 from http://www.icad.org.

Kahneman, D., and Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, 39, 341-350.

Keller, K., and Colucci, D. (1998). Perception and HMDs: What is it in head-mounted displays that really makes them so terrible? *Helmet- and Head-Mounted Displays III, Proceedings of SPIE*, 3362, 46-53.

Kramer, G. (1994). An introduction to auditory display. In: Kramer, G. (Ed.), *Auditory Display: Sonification, Audification, and Auditory Interfaces.* Reading, MA: Addison Wesley.

Logvinenko, A.D., Adelson, E.H., Ross, D.A., and Somers, D.C. (2005). Straightness as a cue for luminance edge classification. *Perception and Psychophysics*, 67, 120-128.

MacDonald, J., Henry, P., and Letowski, T. (2006). Spatial audio: A bone conduction interface. *International Journal of Audiology*, 45, 595-599.

National Research Council (1997). *Tactical display for soldiers: Human factors considerations*. Washington, DC: National Academy Press.

Rash, C.E. (Editor). (2001). *Helmet-mounted displays: Design issues for rotary-wing aircraft*. Bellingham, WA: SPIE Press.

Rash, C.E. (2008). Information obtained via interviews with AH-64 Apache pilots.

Rash, C.E., and Verona, R.W. (1992). The human factor considerations of image intensification and thermal imaging systems. In: Karim, M.A. (Editor), *Electro-Optical Displays*. New York: Marcel Dekker.

Sibley, F. (1964). Book review: *Perception and the Physical World*, *The Philosophical Review*, 73, 404-408.

Tannen, R.S. (1998). Breaking the sound barrier: Designing auditory displays for global usability. *Proceedings of 4th Conference on Human Factors and the Web*. Basking Ridge, NJ.

Toet, A. (2003). Color the night: applying daytime colors to nighttime imagery. *Enhanced and Synthetic Vision, Proceedings of the SPIE,* 5081, 168-178.

Walkenstein, J.A. (1999). Color night vision: A critical information multiplier. *Proceedings of American Lighting Institute.*

Walker, B., and Stanley, R. (2005). Thresholds of audibility for bone conduction headsets. Presented at the ICAD 05, Eleventh Meeting of *the International Conference on Auditory Display*, Limerick, Ireland, July 6-9.

Walton, G.E., Bower, N.J.A., and Power, T.G.R. (1992). Recognition of familiar faces by newborns. *Infant Behavior and Development*, 15, 265-269.

Warren, W.H. (1998). Heading is a brain in the neck. *Nature Neuroscience*, 1, 647–649.

Wason, P., and Shapiro, D. (1971). Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental Psychology*, 23, 63-71.

Wilkie, R.M., and Wann, J.P. (2003). Eye-movements aid the control of locomotion. *Journal of Vision*, 3, 677-684.